Application No: 10522341   Version No: 3.0

Input Set:

Output Set:

Started:	2010-07-22 15:40:42.953
Finished:	2010-07-22 15:40:50.330
Elapsed:	0 hr(s) 0 min(s) 7 sec(s) 377 ms
Total Warnings:	126
Total Errors:	0
No. of SeqIDs Defined:	179
Actual SeqID Count:	179

Error code	Error Description
W 213	Artificial or Unknown found in <213> in SEQ ID (3)
W 213	Artificial or Unknown found in <213> in SEQ ID (4)
W 402	Undefined organism found in <213> in SEQ ID (13)
W 402	Undefined organism found in <213> in SEQ ID (14)
W 402	Undefined organism found in <213> in SEQ ID (15)
W 402	Undefined organism found in <213> in SEQ ID (16)
W 213	Artificial or Unknown found in <213> in SEQ ID (49)
W 213	Artificial or Unknown found in <213> in SEQ ID (50)
W 213	Artificial or Unknown found in <213> in SEQ ID (51)
W 213	Artificial or Unknown found in <213> in SEQ ID (52)
W 213	Artificial or Unknown found in <213> in SEQ ID (53)
W 213	Artificial or Unknown found in <213> in SEQ ID (54)
W 213	Artificial or Unknown found in <213> in SEQ ID (55)
W 213	Artificial or Unknown found in <213> in SEQ ID (56)
W 213	Artificial or Unknown found in <213> in SEQ ID (57)
W 213	Artificial or Unknown found in <213> in SEQ ID (58)
W 213	Artificial or Unknown found in <213> in SEQ ID (69)
W 213	Artificial or Unknown found in <213> in SEQ ID (70)
W 213	Artificial or Unknown found in <213> in SEQ ID (71)
W 213	Artificial or Unknown found in <213> in SEQ ID (72)

```
          Input Set:


          Output Set:



             Started:   2010-07-22 15:40:42.953
            Finished:   2010-07-22 15:40:50.330
             Elapsed:   0 hr(s) 0 min(s) 7 sec(s) 377 ms
      Total Warnings:   126
        Total Errors:   0
  No. of SeqIDs Defined:   179
   Actual SeqID Count:   179
```

Error code   Error Description

W   213   Artificial or Unknown found in <213> in SEQ ID (73)

W   213   Artificial or Unknown found in <213> in SEQ ID (74)

W   213   Artificial or Unknown found in <213> in SEQ ID (75)

W   213   Artificial or Unknown found in <213> in SEQ ID (76)
                This error has occured more than 20 times, will not be displayed

<210> 1
<211> 1284
<212> DNA
<213> Escherichia coli

<220>
<221> CDS
<222> (1)..(1281)
<223> coding for cytosine deaminase (codA)

<400> 1

```
gtg tcg aat aac gct tta caa aca att att aac gcc cgg tta cca ggc      48
Val Ser Asn Asn Ala Leu Gln Thr Ile Ile Asn Ala Arg Leu Pro Gly
  1               5                  10                  15
gaa gag ggg ctg tgg cag att cat ctg cag gac gga aaa atc agc gcc      96
Glu Glu Gly Leu Trp Gln Ile His Leu Gln Asp Gly Lys Ile Ser Ala
             20                  25                  30
att gat gcg caa tcc ggc gtg atg ccc ata act gaa aac agc ctg gat     144
Ile Asp Ala Gln Ser Gly Val Met Pro Ile Thr Glu Asn Ser Leu Asp
         35                  40                  45
gcc gaa caa ggt tta gtt ata ccg ccg ttt gtg gag cca cat att cac     192
Ala Glu Gln Gly Leu Val Ile Pro Pro Phe Val Glu Pro His Ile His
     50                  55                  60
ctg gac acc acg caa acc gcc gga caa ccg aac tgg aat cag tcc ggc     240
Leu Asp Thr Thr Gln Thr Ala Gly Gln Pro Asn Trp Asn Gln Ser Gly
 65                  70                  75                  80
acg ctg ttt gaa ggc att gaa cgc tgg gcc gag cgc aaa gcg tta tta     288
Thr Leu Phe Glu Gly Ile Glu Arg Trp Ala Glu Arg Lys Ala Leu Leu
                 85                  90                  95
acc cat gac gat gtg aaa caa cgc gca tgg caa acg ctg aaa tgg cag     336
Thr His Asp Asp Val Lys Gln Arg Ala Trp Gln Thr Leu Lys Trp Gln
             100                 105                 110
att gcc aac ggc att cag cat gtg cgt acc cat gtc gat gtt tcg gat     384
Ile Ala Asn Gly Ile Gln His Val Arg Thr His Val Asp Val Ser Asp
         115                 120                 125
gca acg cta act gcg ctg aaa gca atg ctg gaa gtg aag cag gaa gtc     432
Ala Thr Leu Thr Ala Leu Lys Ala Met Leu Glu Val Lys Gln Glu Val
     130                 135                 140
gcg ccg tgg att gat ctg caa atc gtc gcc ttc cct cag gaa ggg att     480
Ala Pro Trp Ile Asp Leu Gln Ile Val Ala Phe Pro Gln Glu Gly Ile
145                 150                 155                 160
ttg tcg tat ccc aac ggt gaa gcg ttg ctg gaa gag gcg tta cgc tta     528
Leu Ser Tyr Pro Asn Gly Glu Ala Leu Leu Glu Glu Ala Leu Arg Leu
                 165                 170                 175
ggg gca gat gta gtg ggg gcg att ccg cat ttt gaa ttt acc cgt gaa     576
Gly Ala Asp Val Val Gly Ala Ile Pro His Phe Glu Phe Thr Arg Glu
             180                 185                 190
tac ggc gtg gag tcg ctg cat aaa acc ttc gcc ctg gcg caa aaa tac     624
Tyr Gly Val Glu Ser Leu His Lys Thr Phe Ala Leu Ala Gln Lys Tyr
         195                 200                 205
gac cgt ctc atc gac gtt cac tgt gat gag atc gat gac gag cag tcg     672
Asp Arg Leu Ile Asp Val His Cys Asp Glu Ile Asp Asp Glu Gln Ser
     210                 215                 220
cgc ttt gtc gaa acc gtt gct gcc ctg gcg cac cat gaa ggc atg ggc     720
Arg Phe Val Glu Thr Val Ala Ala Leu Ala His His Glu Gly Met Gly
225                 230                 235                 240
gcg cga gtc acc gcc agc cac acc acg gca atg cac tcc tat aac ggg     768
```

```
Ala Arg Val Thr Ala Ser His Thr Thr Ala Met His Ser Tyr Asn Gly
                245                 250                 255
gcg tat acc tca cgc ctg ttc cgc ttg ctg aaa atg tcc ggt att aac   816
Ala Tyr Thr Ser Arg Leu Phe Arg Leu Leu Lys Met Ser Gly Ile Asn
            260                 265                 270
ttt gtc gcc aac ccg ctg gtc aat att cat ctg caa gga cgt ttc gat   864
Phe Val Ala Asn Pro Leu Val Asn Ile His Leu Gln Gly Arg Phe Asp
        275                 280                 285
acg tat cca aaa cgt cgc ggc atc acg cgc gtt aaa gag atg ctg gag   912
Thr Tyr Pro Lys Arg Arg Gly Ile Thr Arg Val Lys Glu Met Leu Glu
    290                 295                 300
tcc ggc att aac gtc tgc ttt ggt cac gat gat gtc ttc gat ccg tgg   960
Ser Gly Ile Asn Val Cys Phe Gly His Asp Asp Val Phe Asp Pro Trp
305                 310                 315                 320
tat ccg ctg gga acg gcg aat atg ctg caa gtg ctg cat atg ggg ctg  1008
Tyr Pro Leu Gly Thr Ala Asn Met Leu Gln Val Leu His Met Gly Leu
                325                 330                 335
cat gtt tgc cag ttg atg ggc tac ggg cag att aac gat ggc ctg aat  1056
His Val Cys Gln Leu Met Gly Tyr Gly Gln Ile Asn Asp Gly Leu Asn
            340                 345                 350
tta atc acc cac cac agc gca agg acg ttg aat ttg cag gat tac ggc  1104
Leu Ile Thr His His Ser Ala Arg Thr Leu Asn Leu Gln Asp Tyr Gly
        355                 360                 365
att gcc gcc gga aac agc gcc aac ctg att atc ctg ccg gct gaa aat  1152
Ile Ala Ala Gly Asn Ser Ala Asn Leu Ile Ile Leu Pro Ala Glu Asn
    370                 375                 380
ggg ttt gat gcg ctg cgc cgt cag gtt ccg gta cgt tat tcg gta cgt  1200
Gly Phe Asp Ala Leu Arg Arg Gln Val Pro Val Arg Tyr Ser Val Arg
385                 390                 395                 400
ggc ggc aag gtg att gcc agc aca caa ccg gca caa acc acc gta tat  1248
Gly Gly Lys Val Ile Ala Ser Thr Gln Pro Ala Gln Thr Thr Val Tyr
                405                 410                 415
ctg gag cag cca gaa gcc atc gat tac aaa cgt tga                  1284
Leu Glu Gln Pro Glu Ala Ile Asp Tyr Lys Arg
                420                 425


<210> 2
<211> 427
<212> PRT
<213> Escherichia coli

<400> 2
Val Ser Asn Asn Ala Leu Gln Thr Ile Ile Asn Ala Arg Leu Pro Gly
  1               5                   10                  15
Glu Glu Gly Leu Trp Gln Ile His Leu Gln Asp Gly Lys Ile Ser Ala
            20                  25                  30
Ile Asp Ala Gln Ser Gly Val Met Pro Ile Thr Glu Asn Ser Leu Asp
        35                  40                  45
Ala Glu Gln Gly Leu Val Ile Pro Pro Phe Val Glu Pro His Ile His
    50                  55                  60
Leu Asp Thr Thr Gln Thr Ala Gly Gln Pro Asn Trp Asn Gln Ser Gly
 65                 70                  75                  80
Thr Leu Phe Glu Gly Ile Glu Arg Trp Ala Glu Arg Lys Ala Leu Leu
                85                  90                  95
Thr His Asp Asp Val Lys Gln Arg Ala Trp Gln Thr Leu Lys Trp Gln
            100                 105                 110
```

```
Ile Ala Asn Gly Ile Gln His Val Arg Thr His Val Asp Val Ser Asp
        115                 120                 125
Ala Thr Leu Thr Ala Leu Lys Ala Met Leu Glu Val Lys Gln Glu Val
    130                 135                 140
Ala Pro Trp Ile Asp Leu Gln Ile Val Ala Phe Pro Gln Glu Gly Ile
145                 150                 155                 160
Leu Ser Tyr Pro Asn Gly Glu Ala Leu Leu Glu Glu Ala Leu Arg Leu
                165                 170                 175
Gly Ala Asp Val Val Gly Ala Ile Pro His Phe Glu Phe Thr Arg Glu
                180                 185                 190
Tyr Gly Val Glu Ser Leu His Lys Thr Phe Ala Leu Ala Gln Lys Tyr
        195                 200                 205
Asp Arg Leu Ile Asp Val His Cys Asp Glu Ile Asp Asp Glu Gln Ser
    210                 215                 220
Arg Phe Val Glu Thr Val Ala Ala Leu Ala His His Glu Gly Met Gly
225                 230                 235                 240
Ala Arg Val Thr Ala Ser His Thr Thr Ala Met His Ser Tyr Asn Gly
                245                 250                 255
Ala Tyr Thr Ser Arg Leu Phe Arg Leu Leu Lys Met Ser Gly Ile Asn
                260                 265                 270
Phe Val Ala Asn Pro Leu Val Asn Ile His Leu Gln Gly Arg Phe Asp
        275                 280                 285
Thr Tyr Pro Lys Arg Arg Gly Ile Thr Arg Val Lys Glu Met Leu Glu
    290                 295                 300
Ser Gly Ile Asn Val Cys Phe Gly His Asp Asp Val Phe Asp Pro Trp
305                 310                 315                 320
Tyr Pro Leu Gly Thr Ala Asn Met Leu Gln Val Leu His Met Gly Leu
                325                 330                 335
His Val Cys Gln Leu Met Gly Tyr Gly Gln Ile Asn Asp Gly Leu Asn
        340                 345                 350
Leu Ile Thr His His Ser Ala Arg Thr Leu Asn Leu Gln Asp Tyr Gly
        355                 360                 365
Ile Ala Ala Gly Asn Ser Ala Asn Leu Ile Ile Leu Pro Ala Glu Asn
    370                 375                 380
Gly Phe Asp Ala Leu Arg Arg Gln Val Pro Val Arg Tyr Ser Val Arg
385                 390                 395                 400
Gly Gly Lys Val Ile Ala Ser Thr Gln Pro Ala Gln Thr Thr Val Tyr
                405                 410                 415
Leu Glu Gln Pro Glu Ala Ile Asp Tyr Lys Arg
        420                 425
```

<210> 3
<211> 1284
<212> DNA
<213> Artificial sequence

<220>
<223> Description of the artificial sequence: coding for
     cytosine deaminase (codA)

<220>
<221> misc_feature
<222> (1)..(3)
<223> mutation of GTG to ATG start codon for expression
     in eukaryotic hosts

<220>
<221> CDS
<222> (1)..(1281)
<223> coding for cytosine deaminase (codA)

<400> 3
atg tcg aat aac gct tta caa aca att att aac gcc cgg tta cca ggc   48
Met Ser Asn Asn Ala Leu Gln Thr Ile Ile Asn Ala Arg Leu Pro Gly
  1   5   10   15
gaa gag ggg ctg tgg cag att cat ctg cag gac gga aaa atc agc gcc   96
Glu Glu Gly Leu Trp Gln Ile His Leu Gln Asp Gly Lys Ile Ser Ala
                20   25   30
att gat gcg caa tcc ggc gtg atg ccc ata act gaa aac agc ctg gat   144
Ile Asp Ala Gln Ser Gly Val Met Pro Ile Thr Glu Asn Ser Leu Asp
            35   40   45
gcc gaa caa ggt tta gtt ata ccg ccg ttt gtg gag cca cat att cac   192
Ala Glu Gln Gly Leu Val Ile Pro Pro Phe Val Glu Pro His Ile His
        50   55   60
ctg gac acc acg caa acc gcc gga caa ccg aac tgg aat cag tcc ggc   240
Leu Asp Thr Thr Gln Thr Ala Gly Gln Pro Asn Trp Asn Gln Ser Gly
 65   70   75   80
acg ctg ttt gaa ggc att gaa cgc tgg gcc gag cgc aaa gcg tta tta   288
Thr Leu Phe Glu Gly Ile Glu Arg Trp Ala Glu Arg Lys Ala Leu Leu
                85   90   95
acc cat gac gat gtg aaa caa cgc gca tgg caa acg ctg aaa tgg cag   336
Thr His Asp Asp Val Lys Gln Arg Ala Trp Gln Thr Leu Lys Trp Gln
                100   105   110
att gcc aac ggc att cag cat gtg cgt acc cat gtc gat gtt tcg gat   384
Ile Ala Asn Gly Ile Gln His Val Arg Thr His Val Asp Val Ser Asp
            115   120   125
gca acg cta act gcg ctg aaa gca atg ctg gaa gtg aag cag gaa gtc   432
Ala Thr Leu Thr Ala Leu Lys Ala Met Leu Glu Val Lys Gln Glu Val
        130   135   140
gcg ccg tgg att gat ctg caa atc gtc gcc ttc cct cag gaa ggg att   480
Ala Pro Trp Ile Asp Leu Gln Ile Val Ala Phe Pro Gln Glu Gly Ile
145   150   155   160
ttg tcg tat ccc aac ggt gaa gcg ttg ctg gaa gag gcg tta cgc tta   528
Leu Ser Tyr Pro Asn Gly Glu Ala Leu Leu Glu Glu Ala Leu Arg Leu
                165   170   175
ggg gca gat gta gtg ggg gcg att ccg cat ttt gaa ttt acc cgt gaa   576
Gly Ala Asp Val Val Gly Ala Ile Pro His Phe Glu Phe Thr Arg Glu
                180   185   190
tac ggc gtg gag tcg ctg cat aaa acc ttc gcc ctg gcg caa aaa tac   624
Tyr Gly Val Glu Ser Leu His Lys Thr Phe Ala Leu Ala Gln Lys Tyr
            195   200   205
gac cgt ctc atc gac gtt cac tgt gat gag atc gat gac gag cag tcg   672
Asp Arg Leu Ile Asp Val His Cys Asp Glu Ile Asp Asp Glu Gln Ser
        210   215   220
cgc ttt gtc gaa acc gtt gct gcc ctg gcg cac cat gaa ggc atg ggc   720
Arg Phe Val Glu Thr Val Ala Ala Leu Ala His His Glu Gly Met Gly
225   230   235   240
gcg cga gtc acc gcc agc cac acc acg gca atg cac tcc tat aac ggg   768
Ala Arg Val Thr Ala Ser His Thr Thr Ala Met His Ser Tyr Asn Gly
                245   250   255
gcg tat acc tca cgc ctg ttc cgc ttg ctg aaa atg tcc ggt att aac   816
Ala Tyr Thr Ser Arg Leu Phe Arg Leu Leu Lys Met Ser Gly Ile Asn
                260   265   270

```
ttt gtc gcc aac ccg ctg gtc aat att cat ctg caa gga cgt ttc gat      864
Phe Val Ala Asn Pro Leu Val Asn Ile His Leu Gln Gly Arg Phe Asp
        275                 280                 285
acg tat cca aaa cgt cgc ggc atc acg cgc gtt aaa gag atg ctg gag      912
Thr Tyr Pro Lys Arg Arg Gly Ile Thr Arg Val Lys Glu Met Leu Glu
    290                 295                 300
tcc ggc att aac gtc tgc ttt ggt cac gat gat gtc ttc gat ccg tgg      960
Ser Gly Ile Asn Val Cys Phe Gly His Asp Asp Val Phe Asp Pro Trp
305                 310                 315                 320
tat ccg ctg gga acg gcg aat atg ctg caa gtg ctg cat atg ggg ctg     1008
Tyr Pro Leu Gly Thr Ala Asn Met Leu Gln Val Leu His Met Gly Leu
                325                 330                 335
cat gtt tgc cag ttg atg ggc tac ggg cag att aac gat ggc ctg aat     1056
His Val Cys Gln Leu Met Gly Tyr Gly Gln Ile Asn Asp Gly Leu Asn
                340                 345                 350
tta atc acc cac cac agc gca agg acg ttg aat ttg cag gat tac ggc     1104
Leu Ile Thr His His Ser Ala Arg Thr Leu Asn Leu Gln Asp Tyr Gly
        355                 360                 365
att gcc gcc gga aac agc gcc aac ctg att atc ctg ccg gct gaa aat     1152
Ile Ala Ala Gly Asn Ser Ala Asn Leu Ile Ile Leu Pro Ala Glu Asn
    370                 375                 380
ggg ttt gat gcg ctg cgc cgt cag gtt ccg gta cgt tat tcg gta cgt     1200
Gly Phe Asp Ala Leu Arg Arg Gln Val Pro Val Arg Tyr Ser Val Arg
385                 390                 395                 400
ggc ggc aag gtg att gcc agc aca caa ccg gca caa acc acc gta tat     1248
Gly Gly Lys Val Ile Ala Ser Thr Gln Pro Ala Gln Thr Thr Val Tyr
                405                 410                 415
ctg gag cag cca gaa gcc atc gat tac aaa cgt tga                     1284
Leu Glu Gln Pro Glu Ala Ile Asp Tyr Lys Arg
                420                 425


<210> 4
<211> 427
<212> PRT
<213> Artificial sequence

<220>
<223> Description of the artificial sequence: coding for
     cytosine deaminase (codA)

<400> 4
Met Ser Asn Asn Ala Leu Gln Thr Ile Ile Asn Ala Arg Leu Pro Gly
 1               5                  10                  15
Glu Glu Gly Leu Trp Gln Ile His Leu Gln Asp Gly Lys Ile Ser Ala
                20                  25                  30
Ile Asp Ala Gln Ser Gly Val Met Pro Ile Thr Glu Asn Ser Leu Asp
        35                  40                  45
Ala Glu Gln Gly Leu Val Ile Pro Pro Phe Val Glu Pro His Ile His
    50                  55                  60
Leu Asp Thr Thr Gln Thr Ala Gly Gln Pro Asn Trp Asn Gln Ser Gly
 65              70                  75                  80
Thr Leu Phe Glu Gly Ile Glu Arg Trp Ala Glu Arg Lys Ala Leu Leu
                85                  90                  95
Thr His Asp Asp Val Lys Gln Arg Ala Trp Gln Thr Leu Lys Trp Gln
            100                 105                 110
Ile Ala Asn Gly Ile Gln His Val Arg Thr His Val Asp Val Ser Asp
```

```
          115                    120                    125
Ala Thr Leu Thr Ala Leu Lys Ala Met Leu Glu Val Lys Gln Glu Val
     130                    135                    140
Ala Pro Trp Ile Asp Leu Gln Ile Val Ala Phe Pro Gln Glu Gly Ile
145                    150                    155                    160
Leu Ser Tyr Pro Asn Gly Glu Ala Leu Leu Glu Glu Ala Leu Arg Leu
                165                    170                    175
Gly Ala Asp Val Val Gly Ala Ile Pro His Phe Glu Phe Thr Arg Glu
                180                    185                    190
Tyr Gly Val Glu Ser Leu His Lys Thr Phe Ala Leu Ala Gln Lys Tyr
                195                    200                    205
Asp Arg Leu Ile Asp Val His Cys Asp Glu Ile Asp Asp Glu Gln Ser
     210                    215                    220
Arg Phe Val Glu Thr Val Ala Ala Leu Ala His His Glu Gly Met Gly
225                    230                    235                    240
Ala Arg Val Thr Ala Ser His Thr Thr Ala Met His Ser Tyr Asn Gly
                2
```